Molecules and Cells

The Chromatin Accessibility Landscape of Nonalcoholic Fatty Liver Disease Progression

Byeonggeun Kang[1], Byunghee Kang[2], Tae-Young Roh[2], Rho Hyun Seong[1,*], and Won Kim[3,*]

[1]Department of Biological Sciences, Institute of Molecular Biology & Genetics, Seoul National University, Seoul 08826, Korea, [2]Department of Life Sciences, Pohang University of Science and Technology (POSTECH), Pohang 37673, Korea, [3]Department of Internal Medicine, SMG-SNU Boramae Medical Center, Seoul National University College of Medicine, Seoul 07061, Korea
*Correspondence: rhseong@snu.ac.kr (RHS); drwon1@snu.ac.kr (WK)
https://doi.org/10.14348/molcells.2022.0001
www.molcells.org

The advent of the assay for transposase-accessible chromatin using sequencing (ATAC-seq) has shown great potential as a leading method for analyzing the genome-wide profiling of chromatin accessibility. A comprehensive reference to the ATAC-seq dataset for disease progression is important for understanding the regulatory specificity caused by genetic or epigenetic changes. In this study, we present a genome-wide chromatin accessibility profile of 44 liver samples spanning the full histological spectrum of nonalcoholic fatty liver disease (NAFLD). We analyzed the ATAC-seq signal enrichment, fragment size distribution, and correlation coefficients according to the histological severity of NAFLD (healthy control vs steatosis vs fibrotic nonalcoholic steatohepatitis), demonstrating the high quality of the dataset. Consequently, 112,303 merged regions (genomic regions containing one or multiple overlapping peak regions) were identified. Additionally, we found differentially accessible regions (DARs) and performed transcription factor binding motif enrichment analysis and *de novo* motif analysis to determine new biomarker candidates. These data revealed the gene-regulatory interactions and noncoding factors that can affect NAFLD progression. In summary, our study provides a valuable resource for the human epigenome by applying an advanced approach to facilitate diagnosis and treatment by understanding the non-coding genome of NAFLD.

Keywords: assay for transposase-accessible chromatin using sequencing, biomarker, chromatin accessibility, epigenome analysis, nonalcoholic fatty liver disease

INTRODUCTION

Nonalcoholic fatty liver disease (NAFLD) is a metabolic disease in which fat accumulates in the liver without significant alcohol intake (Cholankeril et al., 2017). NAFLD includes both nonalcoholic fatty liver or simple steatosis, which is a relatively early stage of disease and low risk, and nonalcoholic steatohepatitis (NASH), which is an advanced condition with inflammation and fibrosis (Schuster et al., 2018). Due to the complicated and multifactorial mechanisms underlying the development and progression of NAFLD, our understanding of NAFLD remains insufficient.

Genetic alterations associated with NAFLD may play a role in the pathology thereof. Single nucleotide polymorphisms of *PNPLA3*, *GCKR*, and *TM6SF2* are known genetic changes associated with NAFLD progression (Sliz et al., 2018). Genomic, transcriptomic, and epigenomic analyses of tissue samples from NASH patients have recently been reported and potential biomarkers for NAFLD have also been suggested (Govaere et al., 2020; Oh et al., 2020). Another study also compared the changes in DNA methylation throughout the genome to identify epigenetic markers associated with NAFLD severity (Loomba et al., 2018). These studies led to the

Received 3 January, 2022; revised 5 January, 2022; accepted 10 January, 2022; published online 15 April, 2022

eISSN: 0219-1032

identification of several biomarkers specific to various stages of NAFLD progression. However, they are still insufficient to accurately diagnose the stages of NAFLD progression, which is significantly affected by several environmental factors. Thus, epigenetic analysis of the NAFLD progression is essential for understanding the development and progression of the disease.

In recent years, the assay for transposase-accessible chromatin using sequencing (ATAC-seq) has enabled the genome-wide profiling of chromatin accessibility (Corces et al., 2017). Particularly for small amounts of frozen tissues, ATAC-seq has shown wide applicability; however, it has not yet been conducted for NAFLD. Here, we report a comprehensive epigenomic analysis using ATAC-seq technology across the full histological spectrum ranging from healthy controls to fibrotic NASH in an Asian biopsy-proven NAFLD cohort (Kim et al., 2017; Yoo et al., 2021). This provides a valuable resource that can be used to elucidate transcriptional regulation. This study may also help to understand NAFLD caused by regulatory dysfunction.

MATERIALS AND METHODS

Sample collection

The use of human liver samples was approved by the Institutional Review Board (IRB) of SMG-SNU Boramae Medical Center (IRB No. 16-2014-86). All participants were informed of the study protocol and provided written and signed consent. The analyzed population consisted of 19 male and 25 female study subjects, aged 19-80, who visited SMG-SNU Boramae Medical Center. Liver histology was assessed using the NASH Clinical Research Network (CRN) histological scoring system (Sanyal et al., 2021). NAFLD was defined as the presence of >5% macrovesicular steatosis. We graded steatosis, lobular inflammation, and ballooning according to the NAFLD activity score (Kleiner et al., 2005). NASH was determined based on an overall pattern of histological hepatic injury consisting of steatosis, lobular inflammation, and ballooning according to the criteria of Brunt et al. (1999; 2011). Fibrosis was assessed according to a 5-point scale proposed by Brunt and modified by Kleiner et al. (2005): F0, absence of fibrosis; F1, perisinusoidal or periportal fibrosis; F2, perisinusoidal and portal/periportal fibrosis; F3, bridging fibrosis; and F4, cirrhosis (Kleiner et al., 2005). Then, we categorized all study subjects into healthy control, steatosis, and fibrotic NASH according to the histological spectrum of NAFLD. Liver tissues were obtained from 44 individuals through percutaneous needle liver biopsy. Each sample was quickly frozen in liquid nitrogen and stored until nuclei extraction was performed.

The isolation of nuclei and the transposition for ATAC-seq

The flash-frozen tissues were manually dissociated. Isolated nuclei were quantified using a hemocytometer, and 100,000 nuclei were tagmented as previously described (Buenrostro et al., 2013), with some modifications (Corces et al., 2017) using the enzyme and buffer provided in the Nextera Library Prep Kit (Illumina, USA). The tagmented DNA was then purified using the MinElute PCR Purification Kit (Qiagen, Germany), amplified with 10 cycles of polymerase chain reaction (PCR), and purified using Agencourt AMPure SPRI beads (Beckman Coulter, USA). The resulting material was quantified using the KAPA Library Quantification Kit for Illumina platforms (KAPA Biosystems, USA), and sequenced with PE42 sequencing on the NextSeq 500 sequencer (Illumina).

ATAC-seq and computational analysis

ATAC-seq library construction was processed. All ATAC-seq libraries were sequenced on the Illumina NextSeq 500 with 42 bp paired-end reads. The reads were aligned to hg38 genome using bowtie2-2.3.4.1 (Langmead and Salzberg, 2012) with parameters "-k 4 --end-to--end". The reads with duplicates and aligned to mitochondrial DNA were excluded using SAMtools-1.7 (Li et al., 2009) for further analysis. To correct a bias caused by tn5 transposase, all mapped reads were offset by +4 bp for the + strand and −5 bp for the − strand (Buenrostro et al., 2013). To build bigwig track, genomeCoverageBed and bedGraphToBigWig were used. MACS2-2.1.2 (Zhang et al., 2008) was used to call peaks for each sample with parameters "-q 0.05 --nomodel --shift -100 --extsize 200" and peaks on ENCODE blacklist v2 regions were excluded (Amemiya et al., 2019).

Identification of differentially accessible regions (DARs)

DiffBind (Ross-Innes et al., 2012) was used to select differentially accessible peaks with FDR (false discovery rate) < 0.05. The *de novo* motif analysis was performed with DREME and best matches with JASPAR 2018 CORE motifs (Khan et al., 2018) were found using TOMTOM in MEME-suite version 5.0.5 (Bailey et al., 2015). GREAT (genomic regions enrichment of annotations tool) was used to link sets of genomic regions to putative biological functions based on the functional annotations of the nearby genes (McLean et al., 2010).

Data deposition and code availability

The NGS sequence data generated in this paper have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (https://dataview.ncbi.nlm.nih.gov/object/PRJNA725028?reviewer=535ltoopds44f3nkklr49alha4).

The code used for data analysis is available at https://github.com/sysgenlab/Steatosis.

RESULTS

Chromatin accessibility landscape of NAFLD progression

To determine the chromatin accessibility of NAFLD progression, ATAC-seq was performed on liver biopsy samples (Fig. 1A, Supplementary Tables S1 and S2). Liver tissues were obtained from 44 individuals, consisting of 4 healthy controls, 19 simple steatosis patients, and 21 fibrotic NASH patients, through percutaneous needle liver biopsy. The clinical features are presented in Table 1. The clinical information included sex, age, body mass index (27.6 ± 3.5 kg/m^2, mean ± SD), diabetes mellitus, grade of steatosis, steatosis score, lobular inflammation, portal inflammation, hepatocyte ballooning degeneration, NAFLD activity score, gamma-glutamyltransferase, fasting blood sugar, homeostatic model assessment

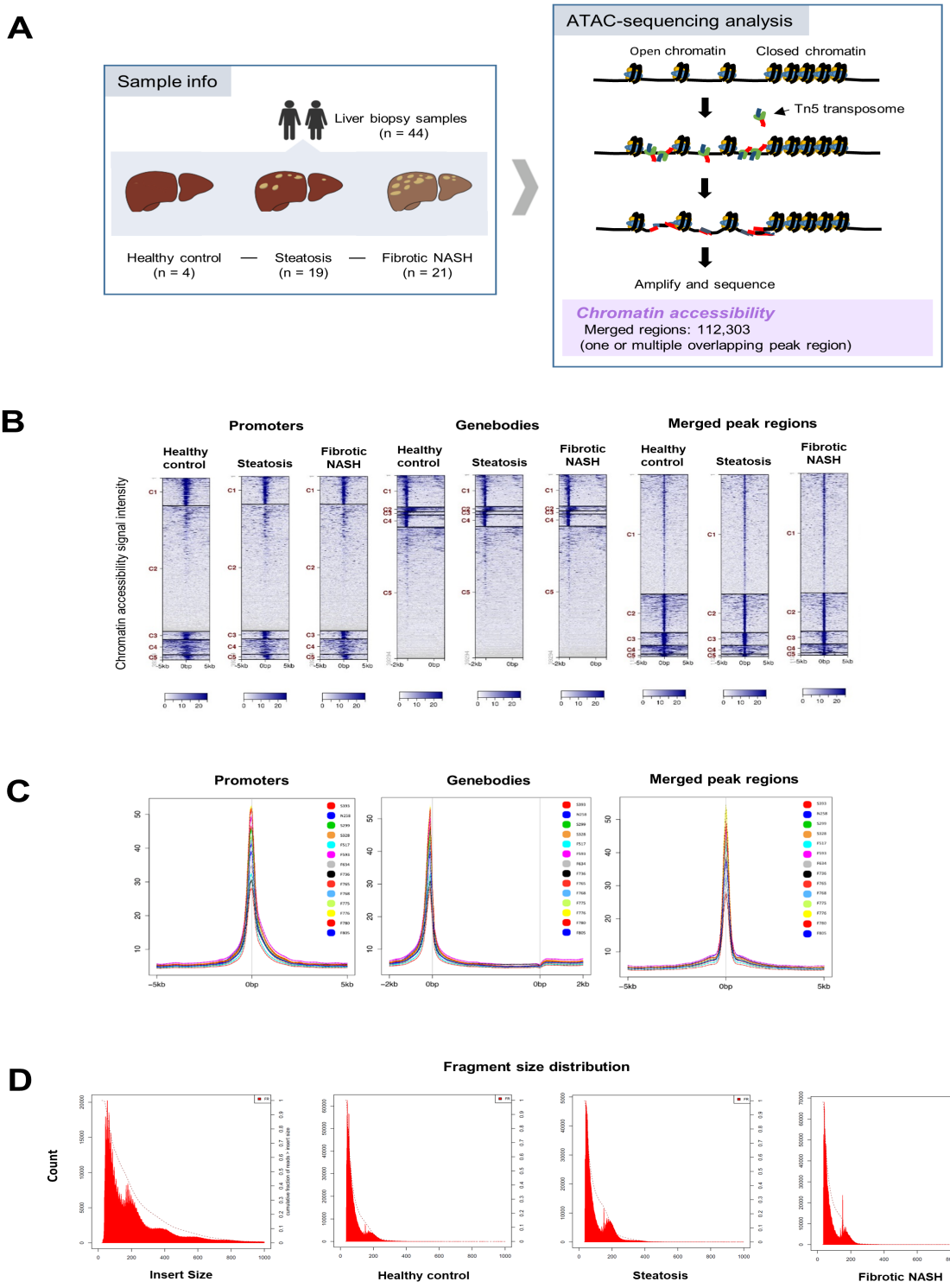


Fig. 1. Chromatin accessibility landscape of NAFLD progression. (A) The workflow overview of the experiment. Forty-four liver samples obtained from the NAFLD study cohort were collected for ATAC-seq profiling. (B) The ATAC-seq signal enrichment around the transcription start sites (promoters), gene bodies and merged peak regions for 3 representative samples (N4, S5, F17). Tag distributions across target regions such as promoters (transcription start sites), gene bodies, or merged peak regions were determined and presented as heatmaps (values in z-axis/color, regions in y-axis). The data were clustered (5 clusters; indicated by C1-C5) and sorted for the heatmaps. (C) The average plot of ATAC-seq profiles for the 14 representative samples (N4, S3, S5, S9, F8, F12-14, F16-21). Tag distributions across target regions such as promoters, gene bodies, or merged peak regions were determined and presented as average plots (average of values for all target regions). (D) The insert size distribution of ATAC-seq profiles for the same samples shown in Fig. 1B.

Table 1. Clinical information

Characteristics	No NAFLD (n = 4)	Steatosis (n = 19)	Fibrotic NASH (n = 21)
Sex, male/female	3/1	10/9	6/15
Age (y)	44.0 ± 12.1	55.3 ± 12.7	53.9 ± 16.4
BMI (kg/m^2)	26.1 ± 3.6	27.1 ± 2.4	28.3 ± 4.1
DM (%)	25.0	42.1	23.8
Grade of steatosis			
<5%	4	-	-
5%-33%	-	13	9
34%-66%	-	5	7
>66%	-	1	5
Steatosis score	0 ± 0	1.4 ± 0.6	1.8 ± 0.8
Lobular inflammation			
Grade 0 (no foci)	1	11	-
Grade 1, <2 foci/200×	3	8	13
Grade 2, 2-4 foci/200×	-	-	7
Grade 3, >4 foci/200×	-	-	1
Portal inflammation			
Grade 0, none	1	12	6
Grade 1, mild	3	7	8
Grade 2, moderate	-	-	6
Grade 3, severe	-	-	1
Ballooning			
Grade 0, none	3	18	2
Grade 1, few	1	1	15
Grade 2, many	-	-	4
NAS	1.0 ± 0.7	1.8 ± 1.0	4.4 ± 1.0
GGT (IU/L)	38.5 ± 21.8	38.2 ± 40.4	68.6 ± 40.8
FBS (mmol/L)	105.5 ± 12.9	103.8 ± 22.8	114.9 ± 26.4
HOMA-IR	2.2 ± 1.0	3.9 ± 1.9	6.2 ± 4.1
Cholesterol (mg/dl)	159.5 ± 17.3	176.1 ± 42.2	176.0 ± 39.4
HDL (mg/dl)	44.3 ± 8.1	42.2 ± 10.5	48.1 ± 10.9
TG (mg/dl)	115.3 ± 28.8	159.0 ± 48.8	148.2 ± 71.7
HbA1c	6.2 ± 1.2	6.1 ± 0.6	6.4 ± 1.3

Data are presented as number only, mean ± SD, or % only.
BMI, body mass index; DM, diabetes mellitus; NAS, NAFLD activity score; GGT, gamma-glutamyltransferase; FBS, fasting blood sugar; HOMA-IR, homeostatic model assessment of insulin resistance; HDL, high-density lipoprotein; TG, triglyceride; HbA1c, hemoglobin A1c (glycated hemoglobin).

of insulin resistance, cholesterol, high-density lipoprotein, triglyceride, and hemoglobin A1c.

Data quality control from the pipeline with the fraction of reads in peaks quality control

We evaluated our ATAC-seq dataset using a series of commonly used statistics, including the number of total aligned pairs, overall alignment rate, de-duplicated pairs, non-mitochondrial, non-redundant fraction, PCR bottleneck coefficient 1, PCR bottleneck coefficient 2, total reads, and the fraction of reads in peaks (FRiP) (Table 2). For each sample, we obtained an average of 50 million reads. The library complexity was measured using the non-redundant fraction, PCR bottleneck coefficient 1, and PCR bottleneck coefficient 2. The successful detection of accessible regions is also supported by the observation of strong enrichment of ATAC-seq reads around transcription start sites (+/– 5 kb, "promoters"), gene bodies (with 2 kb flanking regions), and merged peak regions (equating to all peak regions; +/– 5kb) (Figs. 1B and 1C). ATAC-seq signals are enriched at promoters and merged peak regions, indicating the acceptable quality of the ATAC-seq data. Additionally, the chromatin accessibility fragments showed mainly a mononucleosome peak (Fig. 1D). Together, these results show that high-resolution chromatin accessibility can be obtained from low amounts of frozen tissue samples. We identified accessible regions by using the MACS2-2.1.2 algorithm. We then applied the FRiP quality control and filtered out low-quality data, which was defined as FRiP scores that were less than 20%. Thus, we defined 14 profiles as valid datasets for downstream analysis.

Evaluation of the ATAC-seq data sets

The similarity within each group is shown by heatmap clustering of Pearson correlation coefficients from the comparison of 14 (>20 FRiP) ATAC-seq profiles (Fig. 2A). The correlation heatmap generated by the cross-correlation of two samples based on the read counts of all merged peaks showed a similar pattern. Steatosis and fibrotic NASH were divided into

Table 2. Data QC from the pipeline with FRiP quality control

Sample ID	Status	Total aligned pairs	Overall alignment rate (%)	De-duplicated pairs	Non-mitochondrial	NRF	PBC1	PBC2	Total reads	FRiP (%)
ATAC-seq N1	No NAFLD	34,188,152	87.19	22,598,816	22,063,501	0.947	0.959	25.04	45,197,632	16.99
ATAC-seq N2	No NAFLD	41,351,167	96.6	24,225,839	23,495,342	0.957	0.981	56.773	48,451,678	9.89
ATAC-seq N3	No NAFLD	46,503,438	96.38	27,847,982	27,083,869	0.94	0.961	26.587	55,695,964	10.48
ATAC-seq N4	No NAFLD	33,112,227	88.67	19,673,067	19,043,100	0.948	0.972	38.851	39,346,134	24.85
ATAC-seq S1	Steatosis	30,921,559	84.66	20,852,977	20,371,536	0.973	0.983	61.007	41,705,954	1.6
ATAC-seq S2	Steatosis	31,751,201	88.97	18,743,955	18,106,790	0.941	0.963	28.638	37,487,910	18.37
ATAC-seq S3	Steatosis	29,889,285	96.28	18,669,555	17,933,341	0.945	0.964	29.536	37,339,110	24.97
ATAC-seq S4	Steatosis	28,212,654	84.36	19,515,859	19,045,313	0.97	0.978	47.86	39,031,718	10.51
ATAC-seq S5	Steatosis	37,883,497	91.13	24,394,535	23,733,091	0.941	0.958	24.893	48,789,070	24.21
ATAC-seq S6	Steatosis	56,465,190	96.33	39,159,503	38,577,321	0.961	0.969	32.522	78,319,006	6.67
ATAC-seq S7	Steatosis	50,806,877	92.73	32,286,048	31,576,587	0.949	0.963	27.726	64,572,096	8.58
ATAC-seq S8	Steatosis	38,517,934	88.69	24,053,689	23,355,045	0.947	0.965	30.283	48,107,378	15.73
ATAC-seq S9	Steatosis	37,068,818	97.73	21,180,301	20,297,351	0.936	0.963	29.356	42,360,602	25.88
ATAC-seq S10	Steatosis	45,326,926	96.6	31,777,102	31,327,667	0.974	0.978	46.837	63,554,204	3.88
ATAC-seq S11	Steatosis	39,257,688	97.92	28,486,170	28,231,992	0.976	0.979	47.937	56,972,340	2.33
ATAC-seq S12	Steatosis	41,260,935	97.49	28,987,130	28,590,725	0.97	0.974	39.118	57,974,260	3.01
ATAC-seq S13	Steatosis	39,144,562	98	27,767,522	27,294,895	0.968	0.974	39.576	55,535,044	14.82
ATAC-seq S14	Steatosis	36,726,706	97.95	18,697,086	17,896,692	0.931	0.971	38.762	37,394,172	19.99
ATAC-seq S15	Steatosis	44,258,659	98.01	31,840,113	31,590,256	0.971	0.973	37.66	63,680,226	3.67
ATAC-seq S16	Steatosis	34,627,309	98.25	24,284,265	23,926,454	0.952	0.956	22.637	48,568,530	4.09
ATAC-seq S17	Steatosis	30,146,457	97.58	21,434,866	20,926,433	0.958	0.965	29.857	42,869,732	18.02
ATAC-seq S18	Steatosis	35,623,164	98.06	25,085,724	24,713,880	0.962	0.966	29.939	50,171,448	2.33
ATAC-seq S19	Steatosis	32,202,050	97.69	22,690,629	22,302,683	0.968	0.972	36.622	45,381,258	5.31
ATAC-seq F1	fNASH	41,972,912	90.66	24,999,637	24,295,448	0.944	0.966	30.833	49,999,274	15.14
ATAC-seq F2	fNASH	36,328,466	89.03	24,710,621	24,215,004	0.972	0.981	55.494	49,421,242	5.57
ATAC-seq F3	fNASH	42,112,475	96.45	22,184,474	21,477,269	0.938	0.975	43.797	44,368,948	19.64
ATAC-seq F4	fNASH	33,374,856	97.51	17,310,177	16,593,707	0.941	0.981	61.075	34,620,354	14.48
ATAC-seq F5	fNASH	48,718,809	92.95	30,249,234	29,492,114	0.939	0.955	22.728	60,498,468	12.82
ATAC-seq F6	fNASH	34,441,796	97.94	22,834,839	22,406,271	0.935	0.942	17.397	45,669,678	9.65
ATAC-seq F7	fNASH	50,583,935	98.37	36,928,078	36,545,320	0.964	0.967	31.009	73,856,156	2.75
ATAC-seq F8	fNASH	47,820,895	98.19	33,137,982	32,485,825	0.943	0.952	21.277	66,275,964	20.19
ATAC-seq F9	fNASH	34,913,899	97.6	23,004,148	22,678,142	0.963	0.97	34.621	46,008,296	12.16
ATAC-seq F10	fNASH	41,189,836	97.75	25,680,815	25,479,019	0.943	0.948	19.302	51,361,630	16.23
ATAC-seq F11	fNASH	33,807,906	97.47	20,993,355	20,748,200	0.954	0.963	27.351	41,986,710	9.88
ATAC-seq F12	fNASH	35,340,196	98.11	16,697,422	16,261,806	0.903	0.956	25.029	33,394,844	31.75
ATAC-seq F13	fNASH	41,494,137	98.09	22,157,186	21,740,642	0.926	0.96	27.172	44,314,372	28.73
ATAC-seq F14	fNASH	31,988,959	97.36	19,742,110	19,453,355	0.95	0.965	29.742	39,484,220	26.48
ATAC-seq F15	fNASH	33,261,627	98.36	21,058,448	20,494,281	0.953	0.971	37.313	42,116,896	19.14
ATAC-seq F16	fNASH	59,822,580	98.14	39,692,097	39,079,020	0.942	0.953	21.72	79,384,194	20.38
ATAC-seq F17	fNASH	44,033,686	98.29	27,839,076	27,142,524	0.938	0.959	25.814	55,678,152	26.11
ATAC-seq F18	fNASH	43,324,306	98.26	22,765,064	22,147,927	0.909	0.948	20.417	45,530,128	29.14
ATAC-seq F19	fNASH	42,703,489	98.3	22,160,766	21,517,686	0.906	0.948	21.061	44,321,532	32.76
ATAC-seq F20	fNASH	48,689,030	97.81	30,089,080	29,420,073	0.934	0.956	25.091	60,178,160	30.65
ATAC-seq F21	fNASH	46,063,935	97.09	25,430,585	24,746,270	0.931	0.962	28.434	50,861,170	24.9

De-duplicated pairs: Remove unmapped, not primary alignment, MAPQ < 30 and duplication.
NRF (non-redundant fraction) = distinct read/total reads, PBC1 (PCR bottleneck coefficient 1) = one read/distinct read, PBC2 (PCR bottleneck coefficient 2) = one read/two read. The preferred values are as follows: NRF > 0.9, PBC1 > 0.9, and PBC2 > 3.
FRiP should be >30, though values greater than 20 are acceptable.

separate clusters. This result was also supported by principal component analysis (PCA) (Fig. 2B). PCA classified steatosis and fibrotic NASH along the PC1 axis. The correlation coefficient between samples within the steatosis group was up to 0.97 and the fibrotic NASH group showed a slightly higher correlation coefficient (Figs. 2A and 2C). The high correlation of the ATAC-seq signal between each of the three stages demonstrated high reproducibility. The ability to cluster samples by their disease stage also showed significantly different chromatin accessibility between these three NAFLD stages. The sequencing track of liver fibrogenesis signature genes (*Col1a1*, *Acta2*) (Elbadawy et al., 2020; Seki et al., 2007)

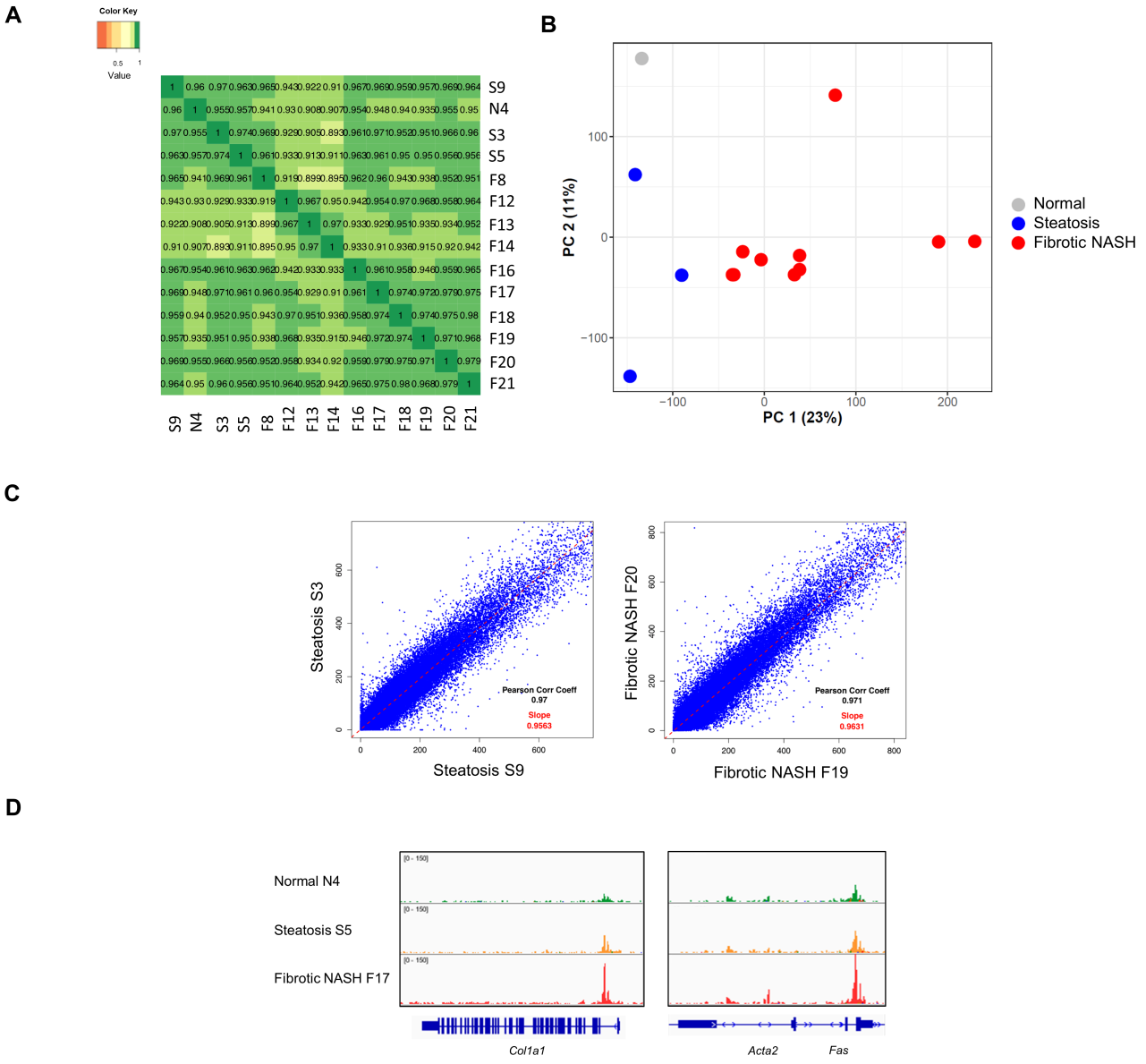


Fig. 2. Evaluation of the ATAC-seq data sets. (A) Heatmap clustering of correlation coefficients across 14 ATAC-seq profiles. This heatmap shows the Pearson Correlation coefficients of all pairwise comparisons. (B) Principal component analysis (PCA) results of 14 samples (>20 FRiP). Each point represents an ATAC-seq sample. Samples with similar chromatin accessibility profiles are clustered together. (C) Peak correlation scatterplot. For each pairwise comparison, scatter plots were generated plotting the tag numbers of sample S9 against S3, F19 against F20 for each merged region. In addition, the slope is a measure for the average ratio in tag numbers between the two samples. (D) Genome browser views of ATAC-seq signal for the indicated fibrosis signature genes (*Col1a1*, *Acta2*).

showed a strong signal in fibrotic NASH samples (Fig. 2D). Thus, these data suggest that our ATAC-seq analyses could detect accessible chromatin regions in the NAFLD patient genome and could be used as a reference dataset for future studies.

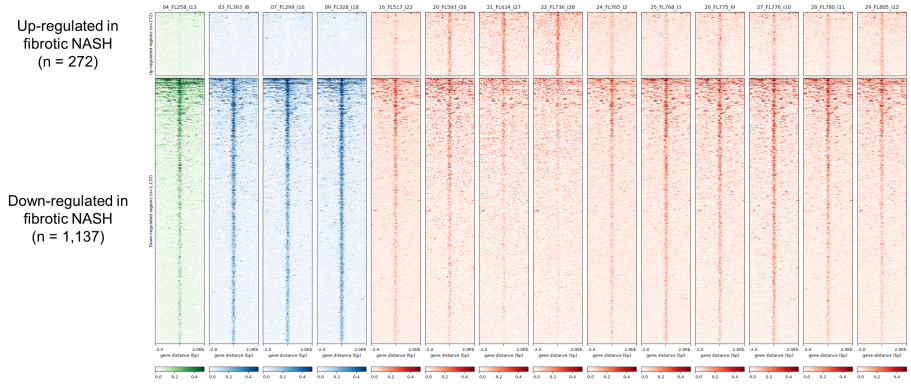
Identification of DARs

To select differentially accessible peaks, DiffBind protocol, originally intended for ChIP-seq data, which constructs a consensus read count matrix from MACS2, was used. When 10 fibrotic NASH samples were compared with 3 steatosis samples, 272 up-regulated and 1,137 down-regulated DARs were identified (Figs. 3A-3C). Functional enrichment analysis was performed on 1,137 down-regulated regions. Significant terms of enriched Gene Ontology were listed in Fig. 3D. Functional enrichment analysis of more closed regions in fibrotic NASH confirmed that these regions were near genes related to hormone response, cell cycle, and liver development. The other DAR results are presented in Supplementary Tables S3-S8.

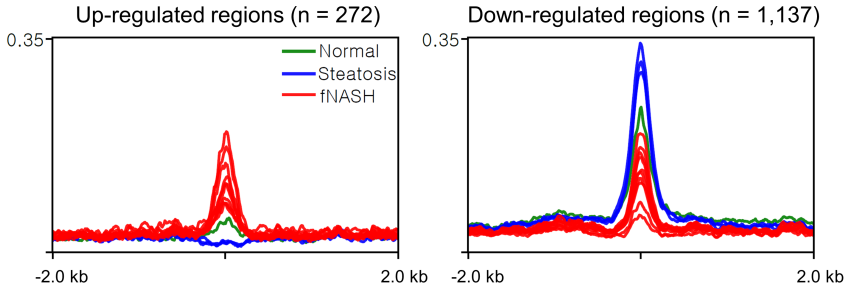
A

Contrasts	Up-regulated	Down-regulated
fNASH vs Steatosis	272	1,137
Steatosis vs Normal	10	39
fNASH vs Normal	10	180

B



C



D

Term	*P* value	FDR
response to peptide hormone	5.41E-09	8.90E-06
cellular response to peptide hormone stimulus	3.74E-08	3.51E-05
positive regulation of G2/M transition of mitotic cell cycle	1.18E-07	7.05E-05
positive regulation of cell cycle G2/M phase transition	1.98E-07	1.00E-04
steroid metabolic process	5.83E-07	2.02E-04
response to cAMP	3.22E-06	6.83E-04
response to ketone	5.25E-06	1.03E-03
cardiac muscle tissue growth	7.91E-06	1.43E-03
hepaticobiliary system development	8.18E-06	1.45E-03
heart growth	8.56E-06	1.46E-03
positive regulation of early endosome to late endosome transport	3.91E-05	4.77E-03
regulation of cell junction assembly	5.05E-05	5.68E-03
liver development	5.39E-05	5.82E-03
positive regulation of heart growth	8.44E-05	8.04E-03
positive regulation of cytoplasmic transport	1.05E-04	8.89E-03
positive regulation of mitotic cell cycle	1.42E-04	1.15E-02
regulation of cell-matrix adhesion	1.62E-04	1.28E-02
phosphatidylinositol 3-kinase signaling	1.85E-04	1.40E-02
endoderm development	1.91E-04	1.42E-02
regulation of cholesterol transport	2.06E-04	1.50E-02

Fig. 3. Identification of DARs associated with NAFLD progression. (A) Number of DARs (up-/down-regulated regions) when comparing disease stages one another. Threshold: FDR < 0.05; FDR: false discovery rate (adjusted *P* value, Benjamini-Hochberg procedure). (B) Heatmap of ATAC-seq signal around DARs when comparing fibrotic NASH to steatosis. (C) Average profile of ATAC-seq signal around DARs when comparing fibrotic NASH to steatosis. (D) Functional enrichment analysis. Fourteen regions (1.2%) are not associated with any genes. GO Biological Process (down-regulated in fNASH: 1,137 regions) was shown.

Transcription factor binding motif and *de novo* motif analyses

To further characterize these DARs, we performed the transcription factor binding site enrichment and *de novo* motif analyses on the 272 regions that were up-regulated in fibrotic NASH and the 1,137 regions that were down-regulated in fibrotic NASH when compared to steatosis (Fig. 4). As NAFLD progresses from steatosis to fibrotic NASH, factors such as E2F6, ELF3, RBPJ, ETV4, IKZF1, OSR2, ETV1, ZKSCAN5, GABPA, SPIC, and EWSR1-FLI1 may play an important role in the regions where chromatin is open, and factors such as ETV1, EHF, GABPA, IKZF1, ELF3, ELF5, SPI1, ELF1, ETV4, STAT2, ZNF384, and IKZF1 in the regions where chromatin is closed. These factors could be promising diagnostic biomarkers that accurately discriminate the stages of NAFLD progression.

DISCUSSION

In the present study, ATAC-seq analysis was conducted to define the changes in chromatin accessibility in liver tissues from healthy controls, steatosis, to fibrotic NASH. Our human liver ATAC-seq data generated for different stages of the NAFLD cohort also suggested open chromatin regions were specifically up- or down-regulated in fibrotic NASH tissues compared to steatosis tissues.

Analysis of high-quality ATAC-seq data obtained from biopsy samples of 44 patients with NAFLD revealed 112,303 merged peak regions (Fig. 1, Supplementary Table S1). These chromatin accessibility profiles identified stage-specific DNA regulatory elements that allow the classification of NAFLD subtypes with newly recognized prognostic importance. We also found 1,648 DARs, showing differences in chromatin openness among the three groups of patients (Fig. 3A, Supplementary Tables S3-S8). In addition, we determined 23 new biomarker candidates through transcription factor (TF) binding motif enrichment and *de novo* motif analysis (Fig. 4, Supplementary Tables S9 and S10).

For example, among the candidates, the chromatin accessibility of EWSR1-FLI1 motif-enriched regions was specifically upregulated in fibrotic NASH (Fig. 4B, upper). EWSR1-FLI1 has an enhancer-regulating role in Ewing sarcoma through divergent chromatin remodeling mechanisms (Riggi et al., 2014). An integrated analysis of chromatin states of Ewing sarcoma showed that EWS-FLI1 could function as a pioneer factor in generating enhancers *de novo* at repeat elements or inhibiting conserved enhancers by competing with endogenous erythroblast transformation specific (ETS) factors. The ETS transcription factor family might be functionally relevant for NAFLD progression since ESE3 overexpression suppressed hepatocellular carcinoma (HCC) (Lyu et al., 2020). ETV4 is positioned in the regulatory module of HCC progression (Kim et al., 2018). The ETS family belongs to candidate TFs whose binding motifs are enriched during the disease progression (Fig. 4A). In addition, SPIC could be a transcription factor that

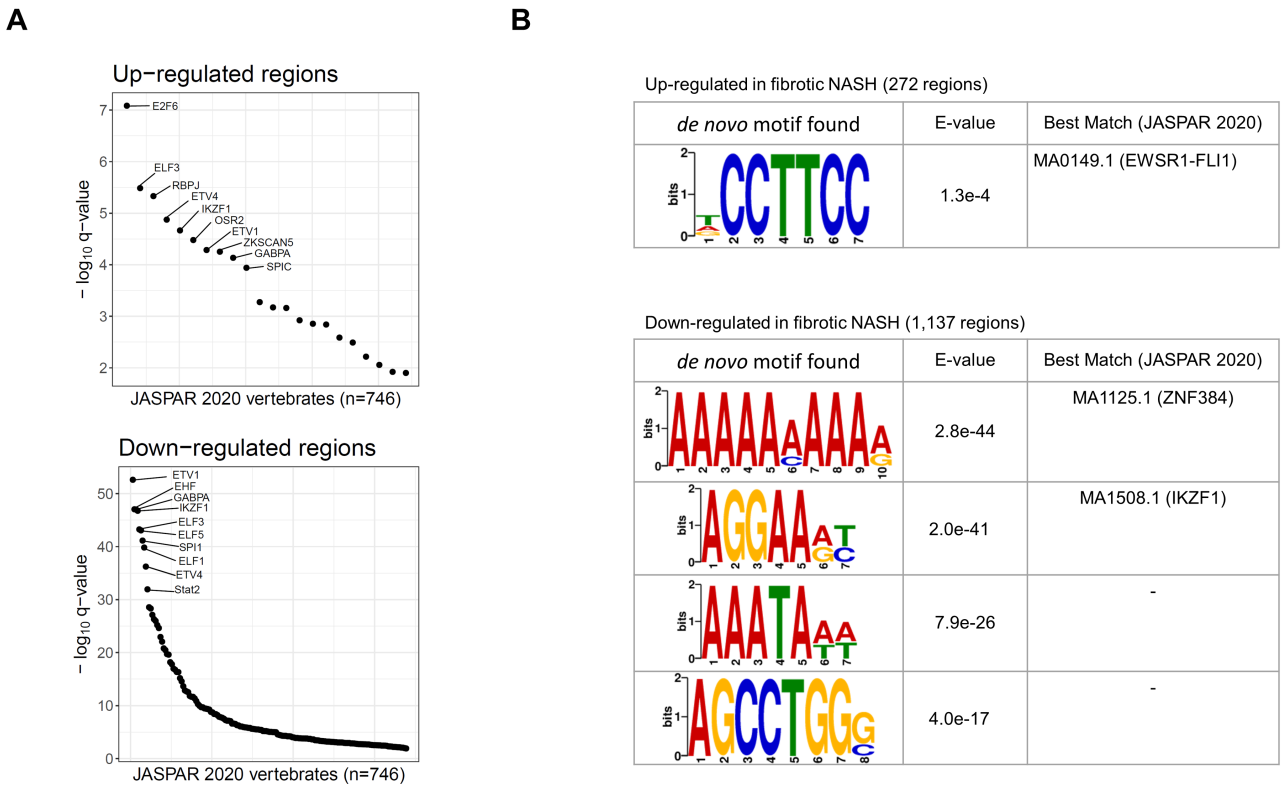


Fig. 4. Transcription factor binding motif analysis and *de novo* motif analysis. (A) Enrichment analysis of JASPAR transcription factor (TF) binding site. The top 10 TF were indicated. (B) Top 1 motif/4 motifs from *de novo* motif analysis at the 272/1,137 differentially accessible chromatin regions which are more open/close in fibrotic NASH when compared to steatosis.

plays a role in the activation of a hepatic stellate cells, which is critical in fibrosis (Marcher et al., 2019). Another attractive candidate is GABPA since it appears in the gene list of both open and closed chromatin. In a previous report by Lake et al. (2016), it was shown that the decreased expression of GABPA may cause down-regulation of its target genes in NASH. It is suggested that these factors could be promising diagnostic and therapeutic biomarkers.

Identifying open chromatin regions is the first step in understanding the regulatory function of non-coding regions, which are not revealed in gene expression changes. However, chromatin accessibility is not always correlated with transcription because histone modifications and DNA methylation status as well as an access to DNA also affect both transcription initiation and its output. Thus, it is necessary to integrate chromatin accessibility with other epigenetic data sets, such as whole-genome bisulfite sequencing and histone modification data (Choi et al., 2020; Gelbart et al., 2009; Shia et al., 2006). The integration of RNA-seq and ATAC-seq will help to find factors whose expression is sufficient for both motif protection and nucleosome positioning. This may enable the development of a quantitative model that links the accessibility of regulatory elements to the expression of the predicted target gene.

In conclusion, the present study will help us understand the epigenetic regulation of the NAFLD progression. Future efforts to combine the genome with the epigenome will pave the way for dealing with the non-coding genome in NAFLD.

Note: Supplementary information is available on the Molecules and Cells website (www.molcells.org).

ACKNOWLEDGMENTS

This research was supported by the Collaborative Genome Program for Fostering New Post-Genome Industry of the National Research Foundation (NRF) and funded by the Ministry of Science and ICT (MIST) (NRF-2017M3C9A6044199) (R.H.S.). This work was supported by the National Research Foundation of Korea grant funded by the Korea government (MEST) (2021R1A2C2005820, 2021M3A9E4021818) (W.K.).

AUTHOR CONTRIBUTIONS

B.K. (Byeonggeun Kang), R.H.S., and W.K. conceived the idea. W.K. collected samples. B.K. (Byeonggeun Kang) generated the data. B.K. (Byeonggeun Kang) and B.K. (Byunghee Kang) analyzed the data. B.K. (Byeonggeun Kang) wrote the manuscript and T.-Y.R., R.H.S., and W.K. supervised the study and revised the manuscript. All authors reviewed and approved the final manuscript.

CONFLICT OF INTEREST

The authors have no potential conflicts of interest to disclose.

ORCID

Byeonggeun Kang https://orcid.org/0000-0003-3269-2745
Byunghee Kang https://orcid.org/0000-0002-6398-5762
Tae-Young Roh https://orcid.org/0000-0001-5833-0844
Rho Hyun Seong https://orcid.org/0000-0001-5699-0718
Won Kim https://orcid.org/0000-0002-2926-1007

REFERENCES

Amemiya, H.M., Kundaje, A., and Boyle, A.P. (2019). The ENCODE blacklist: identification of problematic regions of the genome. Sci. Rep. *9*, 9354.

Bailey, T.L., Johnson, J., Grant, C.E., and Noble, W.S. (2015). The MEME Suite. Nucleic Acids Res. *43*(W1), W39-W49.

Brunt, E.M., Janney, C.G., Di Bisceglie, A.M., Neuschwander-Tetri, B.A., and Bacon, B.R. (1999). Nonalcoholic steatohepatitis: a proposal for grading and staging the histological lesions. Am. J. Gastroenterol. *94*, 2467-2474.

Brunt, E.M., Kleiner, D.E., Wilson, L.A., Belt, P., Neuschwander-Tetri, B.A., and NASH Clinical Research Network (CRN) (2011). Nonalcoholic fatty liver disease (NAFLD) activity score and the histopathologic diagnosis in NAFLD: distinct clinicopathologic meanings. Hepatology *53*, 810-820.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213-1218.

Choi, W.Y., Hwang, J.H., Cho, A.N., Lee, A.J., Jung, I., Cho, S.W., Kim, L.K., and Kim, Y.J. (2020). NEUROD1 intrinsically initiates differentiation of induced pluripotent stem cells into neural progenitor cells. Mol. Cells *43*, 1011-1022.

Cholankeril, G., Patel, R., Khurana, S., and Satapathy, S.K. (2017). Hepatocellular carcinoma in non-alcoholic steatohepatitis: current knowledge and implications for management. World J. Hepatol. *9*, 533-543.

Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods *14*, 959-962.

Elbadawy, M., Yamanaka, M., Goto, Y., Hayashi, K., Tsunedomi, R., Hazama, S., Nagano, H., Yoshida, T., Shibutani, M., Ichikawa, R., et al. (2020). Efficacy of primary liver organoid culture from different stages of non-alcoholic steatohepatitis (NASH) mouse model. Biomaterials *237*, 119823.

Gelbart, M.E., Larschan, E., Peng, S., Park, P.J., and Kuroda, M.I. (2009). Drosophila MSL complex globally acetylates H4K16 on the male X chromosome for dosage compensation. Nat. Struct. Mol. Biol. *16*, 825-832.

Govaere, O., Cockell, S., Tiniakos, D., Queen, R., Younes, R., Vacca, M., Alexander, L., Ravaioli, F., Palmer, J., Petta, S., et al. (2020). Transcriptomic profiling across the nonalcoholic fatty liver disease spectrum reveals gene signatures for steatohepatitis and fibrosis. Sci. Transl. Med. *12*, eaba4448.

Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J.A., van der Lee, R., Bessy, A., Chèneby, J., Kulkarni, S.R., Tan, G., et al. (2018). JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res. *46*(D1), D260-D266.

Kim, E., Kim, D., Lee, J.S., Yoe, J., Park, J., Kim, C.J., Jeong, D., Kim, S., and Lee, Y. (2018). Capicua suppresses hepatocellular carcinoma progression by controlling the ETV4-MMP1 axis. Hepatology *67*, 2287-2301.

Kim, J.Y., Park, K.J., Hwang, J.Y., Kim, G.H., Lee, D., Lee, Y.J., Song, E.H., Yoo, M.G., Kim, B.J., Suh, Y.H., et al. (2017). Activating transcription factor 3 is a target molecule linking hepatic steatosis to impaired glucose homeostasis. J. Hepatol. *67*, 349-359.

Kleiner, D.E., Brunt, E.M., Van Natta, M., Behling, C., Contos, M.J., Cummings, O.W., Ferrell, L.D., Liu, Y.C., Torbenson, M.S., Unalp-Arida, A., et al. (2005). Design and validation of a histological scoring system for nonalcoholic fatty liver disease. Hepatology *41*, 1313-1321.

Lake, A.D., Chaput, A.L., Novak, P., Cherrington, N.J., and Smith, C.L. (2016). Transcription factor binding site enrichment analysis predicts drivers of altered gene expression in nonalcoholic steatohepatitis. Biochem.

Pharmacol. *122*, 62-71.

Langmead, B. and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357-359.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.

Loomba, R., Gindin, Y., Jiang, Z., Lawitz, E., Caldwell, S., Djedjos, C.S., Xu, R., Chung, C., Myers, R.P., Subramanian, G.M., et al. (2018). DNA methylation signatures reflect aging in patients with nonalcoholic steatohepatitis. JCI Insight *3*, e96685.

Lyu, Z., Ma, M., Xu, Y., Wang, X., Zhu, Y., Ren, W., and Li, T. (2020). Expression and prognostic significance of epithelial tissue-specific transcription factor ESE3 in hepatocellular carcinoma. Int. J. Clin. Oncol. *25*, 1334-1345.

Marcher, A.B., Bendixen, S.M., Terkelsen, M.K., Hohmann, S.S., Hansen, M.H., Larsen, B.D., Mandrup, S., Dimke, H., Detlefsen, S., and Ravnskjaer, K. (2019). Transcriptional regulation of Hepatic Stellate Cell activation in NASH. Sci. Rep. *9*, 2324.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495-501.

Oh, S., Jo, Y., Jung, S., Yoon, S., and Yoo, K.H. (2020). From genome sequencing to the discovery of potential biomarkers in liver disease. BMB Rep. *53*, 299-310.

Riggi, N., Knoechel, B., Gillespie, S.M., Rheinbay, E., Boulay, G., Suvà, M.L., Rossetti, N.E., Boonseng, W.E., Oksuz, O., Cook, E.B., et al. (2014). EWS-FLI1 utilizes divergent chromatin remodeling mechanisms to directly activate or repress enhancer elements in Ewing sarcoma. Cancer Cell *26*, 668-681.

Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature *481*, 389-393.

Sanyal, A.J., Van Natta, M.L., Clark, J., Neuschwander-Tetri, B.A., Diehl, A., Dasarathy, S., Loomba, R., Chalasani, N., Kowdley, K., Hameed, B., et al. (2021). Prospective study of outcomes in adults with nonalcoholic fatty liver disease. N. Engl. J. Med. *385*, 1559-1569.

Schuster, S., Cabrera, D., Arrese, M., and Feldstein, A.E. (2018). Triggering and resolution of inflammation in NASH. Nat. Rev. Gastroenterol. Hepatol. *15*, 349-364.

Seki, E., De Minicis, S., Osterreicher, C.H., Kluwe, J., Osawa, Y., Brenner, D.A., and Schwabe, R.F. (2007). TLR4 enhances TGF-beta signaling and hepatic fibrosis. Nat. Med. *13*, 1324-1332.

Shia, W.J., Pattenden, S.G., and Workman, J.L. (2006). Histone H4 lysine 16 acetylation breaks the genome's silence. Genome Biol. *7*, 217.

Sliz, E., Sebert, S., Würtz, P., Kangas, A.J., Soininen, P., Lehtimäki, T., Kähönen, M., Viikari, J., Männikkö, M., Ala-Korpela, M., et al. (2018). NAFLD risk alleles in PNPLA3, TM6SF2, GCKR and LYPLAL1 show divergent metabolic effects. Hum. Mol. Genet. *27*, 2214-2223.

Yoo, T., Joo, S.K., Kim, H.J., Kim, H.Y., Sim, H., Lee, J., Kim, H.H., Jung, S., Lee, Y., Jamialahmadi, O., et al. (2021). Disease-specific eQTL screening reveals an anti-fibrotic effect of AGXT2 in non-alcoholic fatty liver disease. J. Hepatol. *75*, 514-523.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.